Optimal regularized low rank inverse approximation

Julianne Chung\textsuperscript{a,1}, Matthias Chung\textsuperscript{a}, Dianne P. O’Leary\textsuperscript{b,*,2}

\textsuperscript{a} Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA
\textsuperscript{b} Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA

\textbf{ARTICLE INFO}

\textit{Article history:}
Received 5 September 2013
Accepted 19 July 2014
Available online 5 August 2014
Submitted by C. Greif

\textit{MSC:}
65F22
15A29

\textit{Keywords:}
Low rank approximation
Ill-posed problems
Regularization
TSVD
Tikhonov regularization
Inverse problems
Eckart–Young theorem
Approximate pseudoinverse

\textit{ABSTRACT}

In this paper, we consider the problem of finding approximate inverses of a given matrix. We give an explicit solution to the rank-constrained regularized inverse approximation problem and obtain an inverse-regularized Eckart–Young-like theorem.

Published by Elsevier Inc.

\textsuperscript{1} The work of this author was supported by National Science Foundation grant NSF DMS 0902322.

\textsuperscript{2} The work of this author was supported by National Science Foundation grant NSF DMS 1016266.

\textsuperscript{*} Corresponding author.
\textit{E-mail addresses:} jmchung@vt.edu (J. Chung), mcchung@vt.edu (M. Chung), oleary@cs.umd.edu (D.P. O’Leary).

\url{http://dx.doi.org/10.1016/j.laa.2014.07.024}
0024-3795/Published by Elsevier Inc.
1. Introduction

Low rank approximations of matrices and their inverses play an important role in many applications, such as machine learning, matrix completion problems, signal processing, optimal control problems, statistics, and mathematical biology [1–6]. Throughout this paper we consider a rank \( r \) matrix \( A \in \mathbb{R}^{m \times n} \), with \( m \geq n \), and we seek an approximation of rank \( \hat{r} \leq r \).

Let \( U \Sigma V^\top \) be the singular value decomposition of \( A \) with orthogonal matrices \( U = [u_1, \ldots, u_m] \in \mathbb{R}^{m \times m} \) and \( V = [v_1, \ldots, v_n] \in \mathbb{R}^{n \times n} \) and \( m \times n \) matrix \( \Sigma \) containing the singular values \( \sigma_1 \geq \ldots \geq \sigma_n \geq 0 \) on its main diagonal, with exactly \( r \) nonzero singular values. If \( \| \cdot \|_F \) denotes the Frobenius norm, then an optimal low rank approximation of \( A \), the solution to

\[
\min_{\text{rank}(A) \leq \hat{r}} \| A - \hat{A} \|_F, \tag{1}
\]

is given by \( \hat{A} = U_{\hat{r}} \Sigma_{\hat{r}} V_{\hat{r}}^\top \) with \( U_{\hat{r}} = [u_1, \ldots, u_{\hat{r}}] \), \( V_{\hat{r}} = [v_1, \ldots, v_{\hat{r}}] \) and \( \Sigma_{\hat{r}} = \text{diag}(\sigma_1, \ldots, \sigma_{\hat{r}}) \), and \( \hat{A} \) is unique if and only if \( \sigma_{\hat{r}} > \sigma_{\hat{r}+1} \). This problem was first studied by Schmidt in 1907 (Schmidt’s Approximation Theorem [7]), and again by Eckart and Young in 1936 [8]. A similar result was obtained for all unitarily invariant norms by Mirsky in 1960 [9]. In the literature, this is referred to as the Eckart–Young theorem [10,11], and (1) minimizes the error in approximating \( A \) by a rank \( \hat{r} \) matrix.

A closely related optimization problem arises when computing solutions to inverse problems [12,13], where one may not be interested in a low-rank approximation of \( A \) itself but rather in a low-rank approximation of the left inverse of \( A \). Consider the problem

\[
\min_{\text{rank}(Z) \leq \hat{r}} \| ZA - I_n \|_F^2, \tag{2}
\]

where \( I_n \in \mathbb{R}^{n \times n} \) is the identity matrix. (The square of the norm is introduced for convenience and does not change the solution to the problem.) A similar problem arises when computing approximate inverse preconditioners, subject to sparsity constraints [14,15]. As we show later (after the proof of Theorem 3), optimization problem (2) has no unique minimizer. In fact, the truncated SVD (TSVD) pseudoinverse matrix \( \hat{A}^\dagger = V_{\hat{r}} \Sigma_{\hat{r}}^{-1} U_{\hat{r}}^\top \), or a matrix of this form constructed using any choice of \( \hat{r} \) nonzero singular values and corresponding left and right singular vectors, is a global minimizer. The ill-posedness of problem (2) motivates the need to use regularization.

In this paper, we discuss a general regularized low-rank inverse approximation, and consider the following problem

\[
\min_{\text{rank}(Z) \leq \hat{r}} \| (ZA - I_n)M \|_F^2 + \alpha^2 \| Z \|_F^2, \tag{3}
\]
where $M \in \mathbb{R}^{n \times n}$ is invertible and $\alpha \in \mathbb{R}$. If we interpret $\alpha^2$ as a Lagrange multiplier for an equality constraint that forces $\|Z\|_F^2$ to have a particular value, we can see how the second term constrains the choice of $Z$. The weight matrix $M$ allows us to emphasize some terms in the Frobenius norm over others. A diagonal $M$, for example, can be used to downweight columns of $A$ that have more uncertainty than others. Problem (3) is especially interesting for computing solutions to ill-posed inverse problems in image and signal processing, since once the optimal $Z$ is determined for a particular recording instrument, we can very quickly apply it to multiple sets of data in order to recover the input images or signals. It also arises in subspace clustering problems, and the problem has also been studied by Yu and Schuurmans [16]. For us, this formulation arose in solving a Bayes risk minimization problem [17,18], where matrix $M$ is determined from the probability distribution of the signal and the parameter $\alpha$ is determined from the variance in the distribution of the noise and of the signal.

The paper is structured as follows. The main result is presented in Section 2 in Theorem 1 where the solution to regularized problem (3) is derived. We also solve the related problem

$$\min_{\text{rank}(Z) \leq \rho} \left\| M( AZ - I_m ) \right\|_F^2 + \alpha^2 \left\| Z \right\|_F^2,$$

(4)

where $M \in \mathbb{R}^{m \times m}$ is invertible and $\alpha \in \mathbb{R}$. In Section 3, the important special case where $M = I_n$ in (3) is discussed in more detail. We write our results for real matrices $A$ and $M$; the generalization to complex matrices is straightforward.

We note that Sondermann [19] and later Friedland and Torokhti [20] solved problems (3) and (4) in the case that $\alpha = 0$. We discuss the relationships between our results and theirs in Section 4.

2. Low rank optimization problems

In this section, we consider the low-rank optimization problem (3) and, under suitable hypotheses, construct the unique global minimizer. The matrix $A$ is of dimension $m \times n$. Matrix $M$ can be any invertible $n \times n$ matrix. A key tool in our analysis is the generalized singular value decomposition of $A$ and $M^{-1}$, given by

$$A = U \Sigma G^{-1} \quad \text{and} \quad M = G S^{-1} V^\top,$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices, $\Sigma \in \mathbb{R}^{m \times n}$ and $S \in \mathbb{R}^{n \times n}$ are nonnegative matrices, nonzero only on their main diagonals, with $\Sigma^\top \Sigma + S^\top S = I_n$, and $G \in \mathbb{R}^{n \times n}$ is a nonsingular matrix [21, pp. 236–237]. In this form, the decomposition requires that $m \geq n$. The following theorem gives the solution to our low-rank approximate inverse problem.

**Theorem 1.** Given a matrix $A \in \mathbb{R}^{m \times n}$ of rank $r \leq n \leq m$ and an invertible matrix $M \in \mathbb{R}^{n \times n}$, let their generalized singular value decomposition be $A = U \Sigma G^{-1}$,
$M = GS^{-1}V^\top$. Let $\alpha$ be a given parameter, nonzero unless $r = n = m$. Let $\hat{r} \leq r$ be a given positive integer. Define $D = \Sigma \Sigma^{-2} \Sigma^\top + \alpha^2 I_m$. Let the symmetric matrix 

$H = GS^{-4} \Sigma^\top D^{-1} \Sigma G^\top$ have eigenvalue decomposition $H = \hat{V} \hat{\Lambda} \hat{V}^\top$ with eigenvalues ordered so that $\lambda_j \geq \lambda_i$ for $j < i \leq n$. Then a global minimizer $\hat{Z} \in \mathbb{R}^{n \times m}$ of the problem

$$
\hat{Z} = \arg \min_{\text{rank}(Z) \leq \hat{r}} \|(ZA - I_n)M\|_F^2 + \alpha^2 \|Z\|_F^2
$$

is

$$
\hat{Z} = \hat{V}_\hat{r} \hat{V}_\hat{r}^\top GS^{-2} \Sigma^\top D^{-1} U^\top,
$$

where $\hat{V}_\hat{r}$ contains the first $\hat{r}$ columns of $\hat{V}$. Moreover this $\hat{Z}$ is the unique global minimizer of (5) if and only if $\lambda_\hat{r} > \lambda_{\hat{r}+1}$.

**Proof.** Without loss of generality, let $Z$ have the representation $Z = XB^\top U$, where $X \in \mathbb{R}^{n \times \hat{r}}$ is full rank and $B \in \mathbb{R}^{\hat{r} \times m}$. Then using properties of the Frobenius norm and the trace, the function to be minimized in (5) can be represented as

$$
f(X, B) = \|XB^\top U \Sigma G^{-1} V^\top - GS^{-1} V^\top\|_F^2 + \alpha^2 \|XB^\top M\|_F^2
$$

$$
= \|XB \Sigma S^{-1} - GS^{-1}\|_F^2 + \alpha^2 \|XB\|_F^2
$$

$$
= \text{trace}(S^{-1} \Sigma^\top B^\top X^\top XB \Sigma S^{-1}) - 2 \text{trace}(S^{-1} G^\top XB \Sigma S^{-1}) + \alpha \text{trace}(B^\top X^\top XB).
$$

Let $D = \Sigma S^{-2} \Sigma^\top + \alpha^2 I_m$. Then setting the derivative of $f$ with respect to $B$ to zero gives

$$
X^\top (XBD - GS^{-2} \Sigma^\top) = 0.
$$

(7)

Two properties of $D$ will be useful to us. First, it is invertible, since under the hypotheses of the theorem, either $\Sigma S^{-2} \Sigma^\top$ is positive semi-definite and $\alpha^2 > 0$, or if $r = n = m$ then $\Sigma S^{-2} \Sigma^\top$ is positive definite. Second, $\Sigma^\top D^\nu \Sigma = \Sigma^2 D^\nu$ for any $\nu$, where $D_1 = S^{-2} \Sigma^2 + \alpha^2 I_n$ is the $n \times n$ leading principal submatrix of $D$ and similarly for $\Sigma_1$.

Since $X$ is full rank, $X^\top X$ is invertible, and from condition (7) we obtain

$$
B = (X^\top X)^{-1} X^\top G S^{-2} \Sigma^\top D^{-1}.
$$

(8)

Now that we have a formula for $B$ as a function of $X$, minimizing the function of $Z$ in (5) is equivalent to minimizing

$$
\hat{f}(X) = \text{trace}(S^{-3} \Sigma^2 D_1^{-1} G^\top X (X^\top X)^{-1} X^\top G S^{-3} \Sigma^2 D_1^{-1})
$$

$$
- 2 \text{trace}(S^{-1} G^\top X (X^\top X)^{-1} X^\top G S^{-3} \Sigma^2 D_1^{-1}) + \text{trace}(S^{-1} G^\top G S^{-1})
$$

$$
+ \text{trace}(S^{-1} G^\top XB \Sigma S^{-1}) - 2 \text{trace}(S^{-1} G^\top XB \Sigma S^{-1})
$$

$$
+ \alpha \text{trace}(B^\top X^\top XB).
$$
\[ + \alpha^2 \text{trace}(D^{-1} \Sigma S^{-2} G^T X (X^T X)^{-1} X^T G S^{-2} \Sigma^T D^{-1}) \]
\[ = \text{trace}((S^{-2} \Sigma^2 + \alpha^2 I_n) D_1^{-1} - 2I_n) (S^{-1} G^T X (X^T X)^{-1} X^T G S^{-3} \Sigma^2 D_1^{-1})) \]
\[ + \text{trace}(S^{-1} G^T G S^{-1}) \]
\[ = \text{trace}(-X (X^T X)^{-1} X^T H) + \text{trace}(S^{-1} G^T G S^{-1}), \]

where \( H = GS^{-4} \Sigma^T D^{-1} \Sigma G^T \). Therefore, minimizers of \( \hat{f}(X) \) are maximizers of \( g(X) = \text{trace}(X(X^T X)^{-1} X^T H) \).

Without loss of generality, let \( X = \hat{V}Q \) where \( \hat{V} \) comes from the eigendecomposition of \( H \), \( Q = [q_1, \ldots, q_\hat{r}] \in \mathbb{R}^{n \times \hat{r}} \), \( Q^T Q = I_\hat{r} \), and \( R \in \mathbb{R}^{\hat{r} \times \hat{r}} \) is invertible. Then the projector \( X(X^T X)^{-1} X^T \) has the representation

\[ X(X^T X)^{-1} X^T = \hat{V}Q R^T \hat{Q} \hat{V}^T \hat{V}Q R^T \hat{V}^T = \hat{V}QQ^T \hat{V}, \]

so

\[ g(X) = \text{trace}(X(X^T X)^{-1} X^T H) \]
\[ = \text{trace}(\hat{V}QQ^T \hat{V}^T \hat{V} \Lambda \hat{V}^T) \]
\[ = \text{trace}(\hat{V}QQ^T \Lambda \hat{V}^T) \]
\[ = \text{trace}(Q^T \Lambda Q). \]

Therefore, the goal is to find the \( Q \) that maximizes \( \hat{g}(Q) = \text{trace}(Q^T \Lambda Q) \). By defining \( \omega_i = \sum_{j=1}^{\hat{r}} q_{ij}^2 \) we see that

\[ \hat{g}(Q) = \text{trace}(Q^T \Lambda Q) = \sum_{j=1}^{\hat{r}} q_{ij}^T \Lambda q_j = \sum_{j=1}^{\hat{r}} \sum_{i=1}^{n} \lambda_i q_{ij}^2 = \sum_{i=1}^{n} \lambda_i \omega_i. \]

Note that \( \sum_{i=1}^{n} \omega_i = \|Q\|_F^2 = \sum_{j=1}^{\hat{r}} \|q_j\|^2 = \hat{r} \) and \( 0 \leq \omega_i \leq 1 \), where the upper bound comes from viewing \( Q \) as a submatrix of an \( n \times n \) orthogonal matrix. The values

\[ \omega_i = 1, \text{ for } \lambda_1, \ldots, \lambda_\hat{r}, \text{ and } \omega_i = 0, \text{ for } \lambda_{\hat{r}+1}, \ldots, \lambda_n, \quad (9) \]

maximize \( \hat{g} \). Thus

\[ Q = \begin{bmatrix} \hat{Q} \\ 0 \end{bmatrix} \quad (10) \]

where \( \hat{Q} \) is any \( \hat{r} \times \hat{r} \) orthogonal matrix. Hence all global minimizers of \( f \) can be written as
\[
\hat{Z} = XBU^T = X(X^TX)^{-1}X^TGS^{-2}\Sigma^TD^{-1}U^T \\
= \hat{V}\begin{bmatrix} \hat{Q} \\ 0 \end{bmatrix}[\hat{Q}^T0] \hat{V}^TGS^{-2}\Sigma^TD^{-1}U^T \\
= \hat{V}\begin{bmatrix} I_r \\ 0 \\ 0 \end{bmatrix} \hat{V}^TGS^{-2}\Sigma^TD^{-1}U^T \\
= \hat{V}_r\hat{V}_r^TGS^{-2}\Sigma^TD^{-1}U^T.
\]

Finally note that all possible choices of \(\hat{Q}\) and \(R\) lead to the same minimizer \(\hat{Z}\). Thus the matrix \(\hat{Z}\) solves (5) and is unique if and only if \(\lambda_r > \lambda_{r+1}\), since this condition makes the choice of \(\hat{V}_r\) unique. \(\square\)

It is worth mentioning that for \(r = n\), the solution is given by \(GS^{-2}\Sigma^TD^{-1}U^T\), and for \(r < n\), Eq. (6) is just an orthogonal projection of the full-rank solution.

We have presented a solution for the optimal low-rank left inverse approximation problem (3), and a similar result holds for the related problem (4).

**Theorem 2.** Given a matrix \(A \in \mathbb{R}^{m \times n}\) of rank \(r \leq n \leq m\) and an invertible matrix \(M \in \mathbb{R}^{n \times m}\), let their generalized singular value decomposition be \(A = G^{-1}\Sigma U^T, M = VS^{-1}G\) where \(G \in \mathbb{R}^{m \times m}\) and where \(V \in \mathbb{R}^{m \times m}\) and \(U \in \mathbb{R}^{n \times n}\) are orthogonal. Let \(\alpha\) be a given parameter, nonzero unless \(r = n\). Let \(r \leq r\) be a given positive integer. Define \(D = \Sigma^T S^{-2}\Sigma + \alpha^2 I_n\). Let the symmetric matrix \(H = G^T S^{-2} \Sigma D^{-1} \Sigma^T S^{-2} G\) have eigenvalue decomposition \(H = \hat{V}\Lambda\hat{V}^T\) with eigenvalues ordered so that \(\lambda_j \geq \lambda_i\) for \(j < i \leq m\). Then a global minimizer \(\hat{Z} \in \mathbb{R}^{n \times m}\) of the problem

\[
\hat{Z} = \arg\min_{\text{rank}(Z) \leq \hat{r}} \|M(AZ - I_m)\|_F^2 + \alpha^2 \|Z\|_F^2
\]

is

\[
\hat{Z} = UD^{-1}S^{-2}G\hat{V}_r\hat{V}_r^T,
\]

where \(\hat{V}_r\) contains the first \(\hat{r}\) columns of \(\hat{V}\). Moreover this \(\hat{Z}\) is the unique global minimizer of (11) if and only if \(\lambda_r > \lambda_{r+1}\).

**Proof.** When \(m = n\), (11) is a special case of problem (5) of Theorem 1, since the Frobenius norm of the matrix \(M(AZ - I_m)\) is the same as that of the matrix \((Z^T A^T - I_n)M^T\). When \(m > n\), the matrices in (5) and (11) are of different size, causing subtle changes in the result; for example, the condition \(r = n \leq m\), instead of \(r = n = m\), now guarantees that \(D\) is full-rank. We omit the proof of this theorem, though, since it is analogous to that of Theorem 1. \(\square\)
3. Optimal low rank solution for $M = I_n$

In this section, we consider the low-rank optimization problem (3) in the case that $M = I_n$. The following theorem summarizes our result.

**Theorem 3.** Given a matrix $A \in \mathbb{R}^{m \times n}$ of rank $r \leq n \leq m$, let its singular value decomposition be $A = U\Sigma V^T$ with the singular values arranged in nonincreasing order. Let $\alpha$ be a given parameter, nonzero if $r < n$. Let $\hat{r} \leq r$ be a given positive integer. Then a global minimizer $\hat{Z} \in \mathbb{R}^{n \times m}$ of the problem

$$\hat{Z} = \arg \min_{\text{rank}(Z) \leq \hat{r}} \|ZA - I_n\|_F^2 + \alpha^2\|Z\|_F^2$$  \hspace{1cm} (12)

is

$$\hat{Z} = V_{\hat{r}}\Psi_{\hat{r}}U_{\hat{r}}^\top,$$

where $V_{\hat{r}}$ contains the first $\hat{r}$ columns of $V$, $U_{\hat{r}}$ contains the first $\hat{r}$ columns of $U$, and $\Psi_{\hat{r}} = \text{diag}(\frac{\sigma_{\hat{r}}}{\sigma_{\hat{r}} + \alpha}, \ldots, \frac{\sigma_{r}}{\sigma_{r} + \alpha})$. Moreover, this $\hat{Z}$ is the unique global minimizer of (12) if and only if $\sigma_{\hat{r}} > \sigma_{\hat{r} + 1}$.

**Proof.** This is a corollary of Theorem 1. The important observations are that in this special case, $G = V$, $S = I_n$, $H = VAV^T$, and $\lambda_i = \sigma_i^2/(\sigma_i^2 + \alpha^2)$. \hfill $\square$

In the following paragraphs we provide some remarks for the case $M = I_n$.

The case $\alpha = 0$. If $\alpha = 0$, then a solution to problem (12) still exists but, as noted in [20], it is not unique if $\hat{r} < \text{rank}(A)$. To see this, let $ZA$ have singular value decomposition $WTT^\top$, with at most $\hat{r}$ nonzero singular values $\gamma_i$, and let $w_i^\top$ and $t_i^\top$ denote rows of $W$ and $T$, each having 2-norm equal to one. Then

$$\|ZA - I_n\|_F^2 = \text{trace}((ZA - I_n)(ZA - I_n)^\top)$$

$$= \text{trace}(WTT^\top W^\top) - 2\text{trace}(WTT^\top) + \text{trace}(I_n)$$

$$= \sum_{i=1}^{n} (\gamma_i^2 - 2\gamma_i w_i^\top t_i) + n$$

$$\geq \sum_{i=1}^{\hat{r}} (\gamma_i^2 - 2\gamma_i) + n.$$  \hspace{1cm} (13)

We minimize this last expression by letting each $\gamma_i$ equal 1, $i = 1, \ldots, \hat{r}$, so we now know that for any choice of $Z$ of rank $\hat{r}$,

$$\|ZA - I_n\|_F^2 \geq n - \hat{r}. \hspace{1cm} (14)$$
Recalling that the SVD of $A$ is $U \Sigma V^T$, we can achieve this lower bound by choosing $Z = V_\hat{r} \Sigma_\hat{r}^{-1} U_\hat{r}^T$, or a matrix of this form constructed using any choice of $\hat{r}$ singular values and corresponding singular vectors. So all of these choices are global minimizers for (12) with the global minimum value $f(Z) = n - \hat{r}$. These are the only global minimizers, since achieving the lower bound requires that we choose $\gamma_i$’s equal to one or zero and $w_i = t_i$ for $\gamma_i = 1$. Regularization ($\alpha \neq 0$) dictates a particular choice of singular values.

**Uses of regularized inverses.** Using the singular value decomposition of $A$, Theorem 3 provides an optimal low rank approximation of the regularized inverse. This can be used to compute a filtered solution $x_{\text{filter}}$ to the problem $Ax = b$, useful in computing discrete approximations to ill-posed problems [12]. Using our approximate inverse from Theorem 3, the filtered solution can be written as

$$x_{\text{filter}} \equiv Zb = \sum_{j=1}^{n} \phi_j \frac{u_j^T b}{\sigma_j} v_j,$$

where the filter factors are given by

$$\phi_j = \begin{cases} \frac{\sigma_j^2}{\sigma_j^2 + \alpha^2}, & \text{for } j \leq \hat{r}, \\ 0, & \text{for } j > \hat{r}. \end{cases}$$

We denote this choice of filter factors as the truncated Tikhonov filter. It is worth noting here that previous work on hybrid approaches that incorporate filtering techniques such as Tikhonov regularization within a Krylov subspace method [22,23] can produce solutions that are low-rank and regularized. However, to the best of our knowledge, Theorem 3 is the first derivation of the optimal low-rank approximation to the regularized inverse.

**Solutions for other norms.** Generalization of Theorem 3 to other unitarily invariant norms may not necessarily lead to practical results. Consider, for example, the 2-norm problem

$$\min_{\text{rank}(Z) \leq \hat{r}} \|ZA - I_n\|_2^2 + \alpha^2 \|Z\|_2^2,$$

(13)

with $\hat{r} < n$. Here, $\|F\|_2$ denotes the unitarily invariant spectral norm, equal to the largest singular value of $F$ or, equivalently, the square root of the largest eigenvalue of $FF^T$. For any $Z$ with rank $\hat{r} < n$, there exists a nonzero vector $p$ such that $p^T Z = 0$. Therefore

$$\frac{p^T (ZA - I_n)(ZA - I_n)^T p}{p^T p} = 1,$$

and this is a lower bound on the largest eigenvalue of $(ZA - I_n)(ZA - I_n)^T$. Therefore, $\|ZA - I_n\|_2^2 \geq 1$ for all choices of $Z$. It is, therefore, easy to see that $Z = 0$ is a global minimizer of (13), unique if $\alpha \neq 0$, so the problem is not interesting.
4. Conclusion

In this paper, we solved several optimization problems defining optimal low-rank approximate inverses of a matrix $A$ that has at least as many rows as columns. We included regularization terms, a weighting matrix $M$, and rank constraints. We obtained computable solutions for

$$\min_{\operatorname{rank}(Z) \leq \hat{r}} \| (ZA - I_n)M \|_F^2 + \alpha^2 \| Z \|_F^2$$

and

$$\min_{\operatorname{rank}(Z) \leq \hat{r}} \| M(AZ - I_m)M \|_F^2 + \alpha^2 \| Z \|_F^2,$$

and we provided conditions for uniqueness.

Sondermann [19] and, later, Friedland and Torokhti [20] solved a related problem,

$$\min_{\operatorname{rank}(X) \leq k} \| M - BXC \|_F^2. \quad (14)$$

In addition, an if and only if condition for uniqueness is stated in [20]. They require $\alpha = 0$ but do not restrict as we do to the case that either $B = I_n$ and $C = AM$, or $C = I_m$ and $B = MA$. Yu and Schuurmans [16] solved (14) with an added term $\alpha^2 \| X \|_2^2$ for general unitarily invariant norms, but only under a very restrictive simultaneous diagonalizability assumption that $M$ is diagonalized by premultiplication by the transpose of the left SVD factor of $B$ and postmultiplication by the right SVD factor of $C$. We note that related problems are still open for $\alpha \neq 0$:

$$\min_{\operatorname{rank}(Z) \leq \hat{r}} \| M(ZA - I_n) \|_F^2 + \alpha^2 \| Z \|_F^2 \quad \text{and} \quad \min_{\operatorname{rank}(Z) \leq \hat{r}} \| (AZ - I_m)M \|_F^2 + \alpha^2 \| Z \|_F^2,$$

as well as the full generalization of (14) to nonzero $\alpha$.

The low-rank inverse approximation problem arises in many applications such as image processing, machine learning, and computer vision. In future work we will use our approximate inverses to solve ill-posed inverse problems in practical applications. Our investigations will include Bayes risk formulations, efficient computations for large scale problems, and variations of the Bayes problem [5,24].

Acknowledgements

We are grateful to G.W. (Pete) Stewart for helpful discussions and an alternate proof of Theorem 3, to an anonymous referee for pointing us to [19], and to another referee for helpful comments.
References